DOCUMENT RESUME

ED 117 868 95 EC 081 041

AUTHOR Leinhardt, Gaea
TITLE A Strategy for Program Evaluation. Report No. 1975-16.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO LRDC-1975-16
PUB DATE 75
NOTE 22p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS Administration; *Evaluation Methods; *General Education; Interdisciplinary Approach; Organization; *Program Evaluation; *Research and Development Centers; Researchers; Success Factors

ABSTRACT

Presented is a paper on a strategy for in-house evaluations in the context of an educational research and development facility. The obstacles in conducting an evaluation of colleagues' programs are discussed, and a strategy is proposed which places the evaluator in the role of a coordinator, facilitator, and synthesizer of several separate research efforts in which colleagues in a variety of disciplines combine energies and interests to assess the program being examined. It is noted that recruitment and leadership are accomplished by persuasion and by assuring payoff to the researchers involved in the form of contributions to their own area of research. Tactics for organization and management are discussed in some detail. Preliminary results from ongoing research projects are seen to suggest that the evaluation effort was facilitated and its value increased by multidisciplinary contributions to the research. (Author/SB)

A STRATEGY FOR PROGRAM EVALUATION

Gaea Leinhardt

Learning Research and Development Center

University of Pittsburgh

1975

The research reported herein was supported by the Learning Research and Development Center, supported in part as a research and development center by funds from the National Institute of Education (NIE), United States Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of NIE, and no official endorsement should be inferred.

Abstract

This paper proposes a strategy for in-house evaluations in the context of an educational research and development facility. The obstacles in conducting an evaluation of colleagues' programs are discussed. The proposed strategy places the evaluator in the role of a coordinator, facilitator, and synthesizer of several separate research efforts in which colleagues in a variety of disciplines combine energies and interests to assess the program being examined. Recruitment and leadership were accomplished by persuasion and by assuring payoff to the researchers involved in the form of contributions to their own area of research. Tactics for organization and management are discussed in some detail. Preliminary results suggest that the evaluation effort was facilitated and its value increased by multidisciplinary contributions to the research.

A STRATEGY FOR PROGRAM EVALUATION

Gaea Leinhardt

Learning Research and Development Center
University of Pittsburgh

In this paper I will describe a strategy for the in-house evaluation of a specific curriculum in the context of an educational research and development center. The particular curriculum which is examined is the Individualized Science (IS) curriculum (Champagne & Klopfer, 1974), the particular research and development center is the Learning Research and Development Center (LRDC, Note 1). The need for a new approach to evaluation in this context derives from the problems inherent in the role of an evaluator faced with making public judgments of the value of colleagues' work and from those residing in the need to produce detailed and convincing evidence for an evaluation. In order to solve these problems, a procedure was developed which places the evaluator in the role of coordinator, facilitator, and synthesizer of a cooperative, interdisciplinary research effort which brings a variety of research approaches to bear on a cluster of problems. Here I will describe the issues which led to the genesis of this procedure and detail the general strategy and tactics which were developed to improve its chances of success. Since the effort is on-going, there are no results in the usual sense. However, the undertaking has led to the initiation of specific research projects whose results will bear on the overall evaluation, and these studies are described.

Problem

The problems which confront a researcher engaged in evaluating the products of colleagues in the same organization fall into essentially two

categories. First, there is a set of problems that relate to conflicts inherent in judging the work of a colleague without the protection of anonymity or organizational authority (Scriven, 1972). Second, there are problems posed by the unique characteristics of the program under consideration, which can appear more complex than they really are due to the myopia induced by such close association. The strategy described in this paper is designed to turn these apparent disadvantages into advantages in the evaluation of educational programs.

As has been documented by Scriven (1972) and others (Thrasher, Franklin, & Kittredge, Note 2), it is often difficult for researchers in the same organization to design and carry out studies which are aimed at evaluating the work of their colleagues, especially when the results of that research are to be publicly consumable. Regardless of whether such research is unbiased and objective, it will inevitably be viewed with suspicion by outsiders if the results are favorable. If the results are unfavorable, a storm of protest will be raised about methodology, sensitivity, insightfulness, and finally, the value of the evaluation.[1] Surviving such a storm is difficult at a distance and debilitating when it takes place in one's own back yard. Inevitably, if there are positive results, the "outside" world will be suspicious, if there are negative results, the "inside" world will be suspicious (see Rossi & Williams, 1972).

Possible solutions to this dilemma have been debated in the literature. The most popular solution is to remove the evaluative process from the scene of developmental effort either by bringing in evaluators who are external to the organization or by placing the evaluators in a position of

[1]See the now classic debate over the evaluation of Head Start between Cicirelli and Campbell and Erlebacher in Hellmuth (1970) and a discussion of it by Williams and Evans in Rossi and Williams (1972).

higher organizational authority than the developers (Thrasher et al., Note 2). Unfortunately, this solution raises more problems than it solves. "Outsiders" pay for their ability to be critical without consequence by a loss of knowledge and by a limitation on their sources of information (Rossi, 1972). One way out of this conundrum is to use the potential for increased information, which the in-house researcher can have, to provide the most convincing evidence for the value of the program. The trade-off is between dubiously credible objectivity and unique accessibility to information. Neutrality is no guarantee of high quality research; neither does the lack of neutrality insure against it. In fact, efforts to preserve so-called neutrality can often lead to mediocrity in research by providing unbiased, "clean," but irrelevant data for evaluation.

Evaluation can aim to be serviceable and waive the normal guidelines for research, or it can aim to be research and seek to be serviceable within that context. The issue is one of degree. Clearly an elegant research report delivered three years after the appropriate deadline is of no use. On the other hand, a service document filled with theoretical or methodological errors does not truly fill the service function; instead, it merely clouds the issues. The advantage of emphasizing the research aspect of evaluation is that it permits the researcher to invoke the usual rules of evidence of research. The evaluator can then assume a stance of supporting or rejecting the value of a particular product, in the same way a researcher favors a particular solution to a problem and supplies evidence to back up the position.

Emphasizing either the research aspects or the service aspects of evaluation leaves untouched additional problems that arise from the fact that each new program presents its own set of problems for the researcher. For example, the LRDC science curriculum is extremely complex, having multiple goals and multiple modes of accomplishing those goals. It is a program which not only focuses on increasing cognitive abilities with respect

to science, but which also attempts to teach affective and attitudinal competencies. Further, the society which examines, purchases, and uses the program does not have clear-cut goals for elementary science education. If the goals are complex or if the consumers have not stated with clarity or consensus what they want from an elementary science program, it is extremely difficult to provide evidence that documents the value of the program.

Unfortunately, much of the work that has been done in the evaluation of educational programs by external evaluators has ignored the issue of providing useful information and has instead focused on a set of side issues, such as experimental versus quasi- or non-experimental designs in education. Unless treatments are randomly assigned to the unit of analysis, truly comparable "control" groups are almost never available in educational settings. Therefore, the ensuing debate over the results of so-called experimental studies tends to shift from an analysis of the program and its effects to an argument over whether Group A was the same as, equivalent to, or different from Group B (Hellmuth, 1970). This observation is not meant to imply a rejection of experimental design for evaluation. It is rather an attempt to keep the real problem in focus and to suggest that valuable information is still obtainable through the application of different techniques.

In summary, several problems beset any effort to conduct evaluative research in education. The role of the evaluator can be a source of tension for the evaluator and the organization. The specific program under investigation always contributes its own set of difficulties. Finally, problems of acceptable and realistic methodologies abound. In the remainder of this paper I will examine one approach to solving some of these problems.

Resources

Before detailing the strategy, it is useful to describe briefly the setting and the available resources in which the strategy was developed.

The Learning Research and Development Center is one of 14 federally funded educational laboratories and centers in the United States. LRDC is devoted to the improvement of education through individualization by adapting the content and pace of instruction to the individual needs of the student (Glaser, 1973). It consists of 23 research projects, the majority of which report directly to one of three Center directors. Each project has between one and three Ph.D level people and a support staff. The Center has attracted a number of professional educators, psychologists, and sociologists, all of whom work on diverse topics which tend to converge in their purpose if not in their approach. For the evaluation task, a subset of that community has been drawn together. The subset consists of curriculum developers, evaluators, and some individuals less directly and less obviously related to evaluation who are engaged in basic research in psychology, anthropology, and educational research.

The other resources available for the evaluation effort consist of three local schools which are using the science curriculum. Each school has at least one science classroom in which the Individualized Science program is being used. Each science room services between five and eleven classes of science. The three schools range in the degree to which the program is implemented and the degree to which the surrounding environment supports individualized education. Two of the settings are developmental public schools[2] with several individualized programs in operation, and the third is a parochial school which has no other formally individualized curricula. Although there is no immediate access to "control" schools,

[2]Developmental public schools are schools with which LRDC has a contractual relationship permitting curriculum development work to be conducted.

our experience has been that limited amounts of research can be conducted at other schools in the Pittsburgh area. Limited research can be carried out when a large-scale effort involving testing and other disruptive intervention would not be acceptable but smaller, short-term (one month) investigations would be.

A Strategy

In order to effectively use the resources available in the Center to provide evidence for the value of the science program and to avoid as many of the drawbacks as possible, the schema represented in Figure 1 (page 11) was developed. The figure diagrammatically illustrates how specific resources were brought to bear in solving the problem of establishing the value of the Individualized Science curriculum.[3] The basic strategy was to bring together the diverse abilities and interests extant in the Center to investigate various aspects of the science curriculum while simultaneously offering researchers an opportunity to contribute in any way they wished to their own problem areas. Initially, the schema was worked out with only vague indications of peripheral specializations. It was not expected that all areas of the science program would be studied, nor was it expected that each area would be studied in equal depth. What was expected was that the collection of studies from differing perspectives would contribute greatly to the available knowledge about the way in which this specific program works. There is, of course, no guarantee that the knowledge thus generated will be totally adequate to meet the needs of educational consumers, but traditional approaches to evaluation cannot guarantee that either.

[3]Obviously, this entire effort depends upon both the cooperation and enthusiasm of the entire Science Project. Without their generous support, none of the studies could have been conducted.

Tactics

To operationalize this strategy, tactics had to be developed to attract researchers from different groups, to provide circumstances in which they could continue to work on their own areas while contributing to the science evaluation, and finally, to develop a setting in which ideas which were not fully formed could be aired and improved upon. The tactics used to implement this strategy fall into roughly two categories: those behaviors associated with initial selection and recruitment of researchers, and those behaviors associated with development and maintenance of a task performing group.

In theory, an optimal selection of disciplines to be represented (i.e., the peripheral specialties) could be determined a priori and people assigned to do specific projects so that unique information would be maximized while overlap would be minimized. In reality, however, while disciplines and research areas could be selected in advance, people could not because LRDC is a semi-hierarchical organization rather than a hierarchical one. Leadership is by persuasion rather than by command. Therefore, researchers who were both available and in a relevant area had to be attracted to the task and then worked with to develop studies which maximized the information each could contribute.

Selection and recruitment. Initially, a list of individuals who might be interested in working on the science evaluation effort was drawn up. Some individuals were included because they were connected with the Science or Evaluation Projects. Others were added because their work or work interests indicated that their inclusion would be mutually beneficial. Still others were recommended or volunteered out of personal interest as the initial meetings took place. Several graduate students were included on the list because it was assumed that the opportunity both to obtain thesis material and to develop pilot studies in a focused problem setting would be attractive to them.

After this list was constructed, individuals were contacted one at a time and the basic problems were discussed. On the one hand, the science program and the problem of evaluating it had to be made intriguing, relevant, and seductive. On the other hand, payoff to the individuals involved in terms of their own basic research interest had to be relatively assured. In other words, it was important for the products of research to contribute to both the evaluation problem and the specific disciplines, as represented by the double-headed arrows in Figure 1 (page 11). It was felt that given access to "natural" settings, many researchers would welcome the chance to put some of their work to applied use. The remaining task was to provide a forum in which proposals for research and initial results could be discussed in a supportive way.

Maintenance. Small group meetings, which included those researchers who were already "on board," were held once every two weeks, while the individual meetings to recruit additional researchers were being conducted. Several decisions were made to increase the probability of the meetings' success. They were to be open to interested researchers, however, an effort was made to keep the proportion of active working participants to observers high. Individuals were encouraged to attend meetings or send representatives to see what they were about, but they were discouraged from continuing to attend if they did not intend to do any research in the area. Ground rules for discussion of research proposals were laid down which encouraged directed, constructive criticism and discouraged criticism for its own sake.

The purpose of the meetings was to discuss one by one each of the proposed or in-process studies. The discussions focused on the relevance, feasibility, and significance of each study. In some cases the discussion pointed out so many problems with a particular proposal that the study was abandoned. In other cases discussions wandered away from the initial task of examining proposals to excursions into what evaluation should or could be.

These wanderings were not altogether purposeless for they helped to clarify what the type and extent of the resultant information would be. Along with the general discussion, specific methods for implementing each study in field settings were sketched out.

There were drawbacks associated with this process. By having researchers from so many areas all trying to work together, problems arose concerning their individual roles. For example, some researchers attempted to become evaluators and entirely abandoned their own disciplinary perspectives rather than merging the two. Further, because the majority of the researchers were neither evaluators nor science educators, problems arose over the need to familiarize everyone with science content, placing a burden on the Science Project staff. Finally, the most serious risk lay in the potential of producing disparate fragmented studies which would have no common theme or insight. However, the majority of these difficulties were overcome by continuous discussion in the group meetings during the first two or three months of initial work.

During the first seven-month period, 13 meetings were held. The first meetings were largely organizational and introductory. In these meetings the evaluators functioned as facilitators; they did not attempt to direct the kind of research to be conducted. As the meetings progressed, a format developed of presenting one or two proposals for research or reports of work in progress per meeting. Each proposal was reviewed and discussed with regard to both its substance and its feasibility. It would be a legitimate criticism of this approach that not enough initial guidance was given, so that discouraging false starts were sometimes made. However, it was felt that the error should be in the direction of too little rather than too much directiveness, as this was consistent with the style of the organization.

Five months after the meetings began, an interesting interruption occurred. Research for Better Schools (RBS) requested a joint meeting

to exchange ideas on the evaluation of the science curriculum since they were engaged in evaluating the IS field test units as these became available. Their evaluation effort was of the more traditional outcome assessment type. The importance of the meeting for us was that it forced our group to summarize exactly what we were doing and why. This, in turn, led to a substantial increase in group solidarity.

In summary, recruitment was by invitation and not coercion. No tasks or roles were assigned, they were selected. The maintenance of interest in the task was accomplished through highly task-oriented and reasonably supportive group meetings.

Results to Date

Figure 1 shows the ten basic studies which currently make up the group of studies to be done in science. The specific title of each study appears in italic type on one side of the double-arrowed line; the primary discipline from which the individual comes is in boldface type on the other side. The area other than science evaluation to which the research contributes appears around the outside periphery of the figure. The role of the evaluators is to help facilitate the research and, after all the studies are complete, to coordinate results into a series of documents that can serve evaluative functions.

Turning to the details of Figure 1, two of the evaluators are interested in validating the model that had recently been proposed for investigating classroom processes (Cooley & Lohnes, in press). One aspect of that model concerns the use of time and its effect on student achievement. They investigated the relationship between the number of minutes spent in science, math, and reading, and student achievement in one of the schools in which the science program existed (Leinhardt & Cooley, Note 3). Preliminary results indicate that the amount of time spent in science, independent of entering abilities, is positively related not only to science achieve-

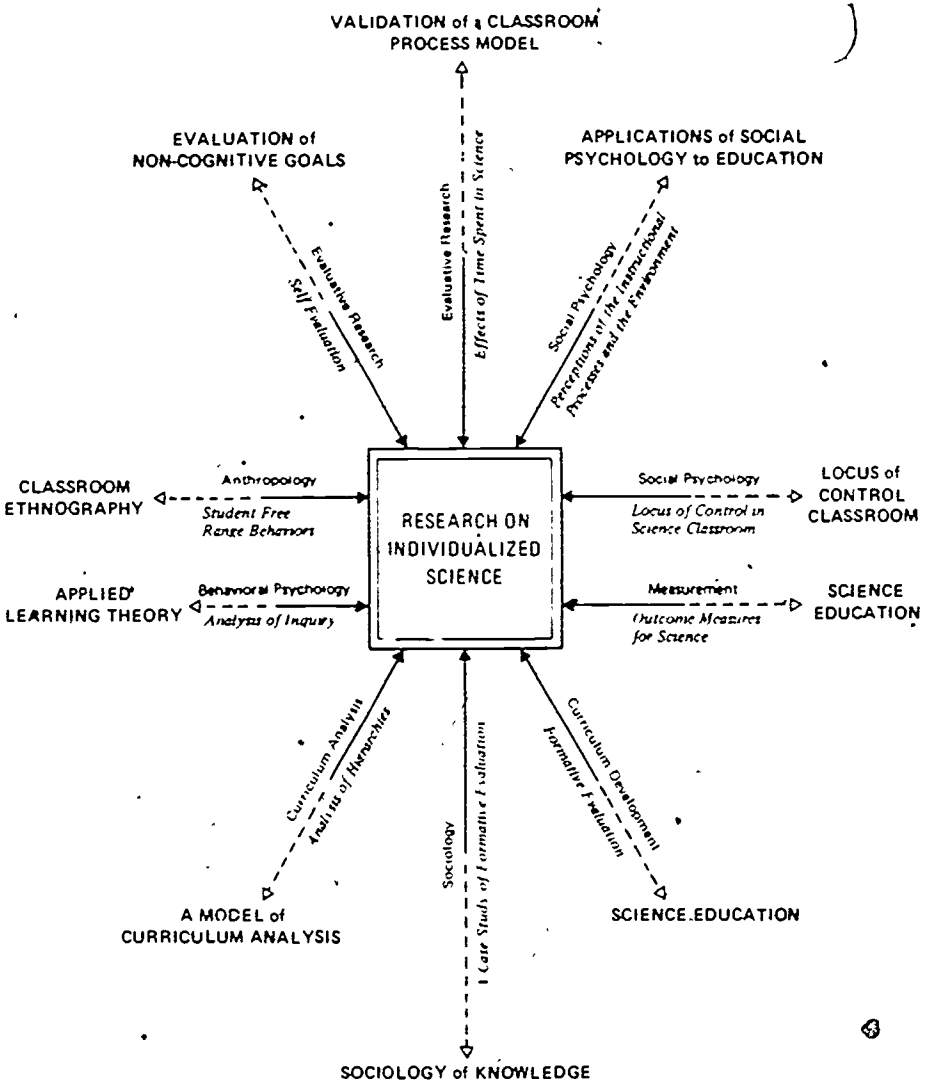

Figure 1 A strategy for science curriculum evaluation

ment, but also to math and reading achievement. This was supported by a similar clinical finding previously reported by teachers and administrators using the science curriculum. This finding has important implications for the value of elementary science education to improve not only the standard of scientific literacy, but also the overall level of knowledge of elementary school children. A replication of this study is currently being conducted.

One of the social psychologists is interested in studying teacher and student perceptions in the classroom. The planning for this research was completed before the meetings were started, however, the meetings were of some help in implementing the plans. This researcher is planning a study in the science rooms to investigate student perceptions of learning, school, and themselves (Greenberg, Note 4). He is examining the students' attitudes and beliefs about knowledge acquisition and the roles of teachers and peers in different environments. This study is being conducted in an inner-city school and will provide information on how children view different learning settings and will shed light on how those views might affect learning itself.

Another social psychologist is interested in the problems of locus of control and of changes in perceptions which occur in different educational environments. He is conducting a study of locus of control within the contexts of the science and mathematics classrooms at two grade levels in a parochial school. One section of mathematics is taught in a traditional and nonindividualized manner, while science is taught in a more open and individualized way. He is working toward the description and measurement of locus of control which is situation specific rather than using the traditional view of it as an underlying trait or characteristic.

The science educators, who have long been concerned with the problems of formative evaluation and its documentation, are working with sociologists who are interested in documenting variations in the frames of reference of actors engaged in formative evaluation. The sociologists are

conducting a case study of the behaviors of the developers during a stage of curriculum development. The product of this joint effort will be a formative evaluation of a specific unit of the science curriculum and a detailed description of how design decisions are made and by whom. Further information about how the different groups involved with dissemination (publishers and school personnel) impact curriculum development will also be documented.

Two researchers are interested in developing methodologies for analyzing curricula. One of the two is interested in developing models of hierarchy analysis. She is detailing the implied and overt hierarchies of both the affective/attitudinal and the scientific literacy goals of the curriculum. The results of this will be a map of the science curriculum's objectives and the specific activities designed to meet these objectives. The second team of researchers is interested in developing measures of how closely curricula use known behavioral principles. These measures have been applied to several diverse curricula (Holland, 1975) and a detailed analysis of one of the Center's other curricula has already been conducted (Holland & Solomon, 1975). They are doing a similar analysis of the science curriculum.

One researcher is doing a small study of the science goal of student self-evaluation. She is interviewing students who have worked through a science unit before they take their posttest for the unit. In the interview the students are asked to estimate how much they know about a given subarea. Their estimation is then compared to their posttest performance. In general, students tend to slightly underestimate their knowledge, with better students being closer in their estimates than poorer students. The importance of this study is that it will help to document the science activities specifically designed to contribute to students' abilities to evaluate themselves, and it will also indicate how other curricula not so designed impede students' ability to evaluate their own knowledge states.

Two other studies have recently emerged in the science meetings. One is an ethnographic description of student free-range behaviors in the science room. The other is the development of new and more extensive science competency measures for elementary and secondary schools (labeled "Measurement" in Figure 1). Currently, most of the standardized tests do not distinguish between science and social studies until the fifth grade. Further, the majority of existing tests are so imprecise and inaccurate in their language as to place a knowledgeable student at a distinct disadvantage.

In summary, these studies contribute to the understanding of IS and hence help to show the value of it along three dimensions. First, reasonably complete documentation of the content, pedagogical soundness, and design decisions of the program will be made available. Second, some insight about specific classroom processes and their impact, both psychological and social, will be generated. Third, the impact of science education on both science and other basic skill areas will be better understood.

Conclusions

An additional result of this exercise has been the validation of the strategy presented in Figure 1. Figure 2 represents a restatement of that strategy independent of the specific studies being conducted. At the center of Figure 2 is the object of the evaluative inquiry. While the object of inquiry is singular, the inquiries are considered plural. That is, there is a set of problems or topics for investigation. Around the periphery of Figure 2 are several problem areas of research, the focus of which is <u>not</u> necessarily evaluative research. These peripheral foci are areas of research which grow out of specific discipline-related inquiries. The figure indicates that the problem areas and related disciplines can be, but are not necessarily, independent of one another. The strategy is to incorporate the work of individuals from diverse backgrounds who are working on problems

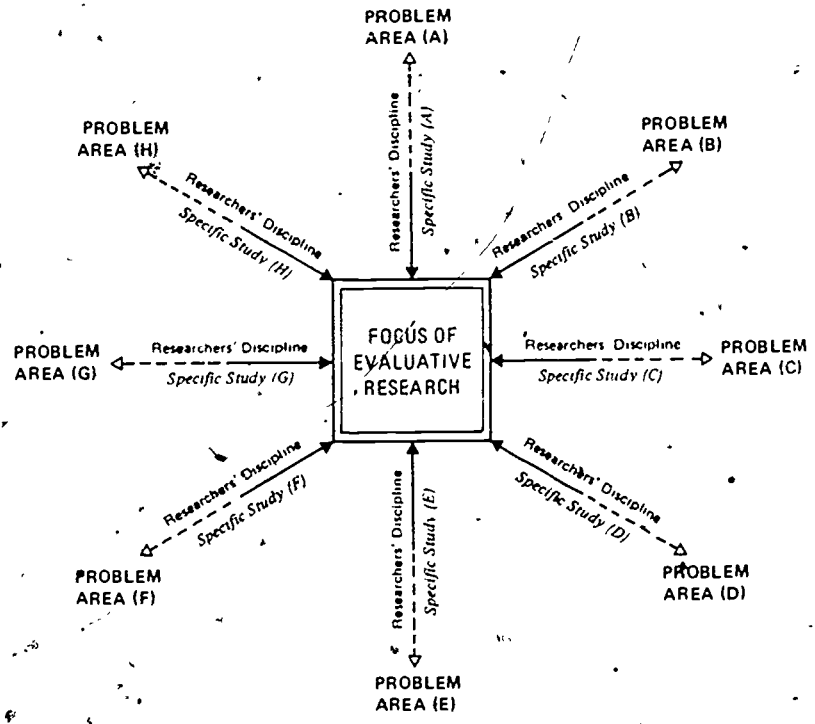

Figure 2 A strategy for program evaluation.

in their own discipline areas into an overall program of evaluative research. These individuals represent a powerful and generally untapped resource for solving many evaluative research problems.

The outcomes of this on-going effort are encouraging. If one wishes, one can exploit the available resources of a research center to facilitate in-house evaluation in a creative fashion resulting in unique and usable information. The role of the evaluator is to orchestrate the undertaking and, when the studies are completed, to incorporate them into an evaluative document. The product of such an effort can be an in-depth view of the program under investigation, one which is richer in expertise and information than any single effort could hope to be.

This strategy was designed to solve several problems: the conflict inherent in the role of in-house evaluator while providing convincing evidence for the value of a program and for the soundness of the evaluation, and the provision of more than superficial information about a highly complex program. The conflict of the role of the evaluator was solved in two ways: first, by sharing the burden of evaluation among several researchers, and second, by including the developers in the group. This assured the developers that misinformation and misinterpretation of information about their program could be minimized. The problem of bias is not totally solved, however, it is reduced by using a group of people whose primary commitment is to good science in their respective disciplines. The problem of providing in-depth information about the program was solved by having multiple, and in some cases, overlapping studies conducted on various facets of the program and by making use of already existing resources, many of which were previously untapped.

An additional significance of this strategy for evaluation should not be overlooked. One of the major complaints of social scientists (such as Rossi) is that the evaluation field has been unable to attract highly qualified individuals and is the refuge of dropouts from other fields (Rossi & Williams,

1972). The strategy described here provides quality researchers from other fields with an incentive to undertake work that serves evaluative functions. Of course, evaluators must do research themselves, but they also must continuously encourage their colleagues who are not evaluators to engage in research that can also serve the function of evaluation. Not only does this provide one with a sense of parsimony, it also educates and re-educates evaluators by exposing them to the differing perspectives and methodologies of their colleagues. A remaining question is whether or not there are other research organizations that have sufficiently diverse personnel to utilize this approach. I think there are. For example, not only are there the university-related research centers, there also are governmental institutes and private industry laboratories, all of which have sufficiently diverse professional staffs to permit utlization of this approach.

The strategy is not offered as a panacea for all the problems that face evaluation. It is offered as an alternative to the traditional in-house, one-shot studies conducted by a single team of evaluators. There is no iron-clad assurance that answers to a collection of questions generated by another group of researchers would provide adequate information about the value of a program. However, the strategy provides ample opportunity for the evaluator and other researchers to discuss this point and try to construct a series of studies that will provide usable information for consumers. The emphasis in this strategy is for researchers to satisfy themselves with regard to the scientific excellence of their work and to seek answers to questions which they generate rather than attempting to second-guess the consumers' questions.

Reference Notes

1. Learning Research and Development Center. Resource allocation and management plan. Research and development for adaptive education. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1973.

2. Thrasher, J. H., Franklin, J. L., & Kittredge, L. D. Evaluation: The internal-external debate reconsidered. Paper presented at the meeting of the American Sociological Association, Montreal, 1974.

3. Leinhardt, G., & Cooley, W. The investigation of the effects of time spent in science on achievement. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, in preparation.

4. Greenberg, M. Research proposal concerning the relationship between social variables and the adaptive classroom. Unpublished manuscript, University of Pittsburgh, 1974.

References

Champagne, A., & Klopfer, L. An individualized elementary school science program. Theory into Practice, 1974, 13, 136-148.

Cooley, W. W., & Lohnes, P. R. Evaluative inquiry in education. New York: Irvington Publishers, in press.

Glaser, R. Educational psychology and education. American Psychologist, 1973, 28, 557-566.

Hellmuth, J. (Ed.). Disadvantaged child. Vol. 3: Compensatory education, a national debate. New York: Brunner/Mazel, 1970.

Holland, J. G. Variables in adaptive decisions in individualized instruction. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1975. (LRDC Publication 1975/10)

Holland, J. G., & Solomon, C. Komputer-konkokted kurrickulum: A review of CAI spelling. Pittsburgh. University of Pittsburgh, Learning Research and Development Center, 1975. (LRDC Publication 1975/15)

Rossi, P. H. Testing for success and failure in social action. In P. H. Rossi & W. Williams (Eds.), Evaluating social programs: Theory practice, and politics. New York: Seminar Press, 1972.

Rossi, P. H., & Williams, W. (Eds.). Evaluating social programs: Theory, practice, and politics. New York: Seminar Press, 1972.

Scriven, M. General strategies in evaluation. In J. Weiss (Ed.), Curriculum evaluation: Potential and reality. Curriculum Theory Network monograph supplement. Toronto, Ontario: Ontario Institute for Studies in Education, 1972.

Williams, W., & Evans, J. The politics of evaluation: The case of Head Start. In P. H. Rossi & W. Williams (Eds.), Evaluating social programs: Theory, practice, and politics. New York: Seminar Press, 1972.